Voice Control

A new way to control your Mac, iPhone and iPad entirely with your voice

September 2019

Contents

Overview

Speech-to-text transcription
The Voice Control speech recognition engine accurately understands and transcribes natural speech, and users can add custom words and commands.

Text editing
With just their voices, users can select text with precision, make fine-grain corrections, and see alternative word and emoji suggestions.

Comprehensive navigation
Users can now access all parts of the screen by saying item names and numbers, using the grid overlay, and recording multistep commands.

Voice gestures
Hand gestures like tap, double tap, and scroll are now voice activated, and users can create customized voice gestures.

Attention awareness
On iPad and iPhone, users can wake up Voice Control and put it to sleep by just looking at and away from their devices.

On-device processing for privacy
Voice Control audio processing happens on-device, so it works online or offline and keeps personal information private.

Voice Control is a new feature built into macOS Catalina, iOS 13, and iPadOS that empowers those who can't use traditional input devices to control their Mac, iPhone, and iPad entirely with their voices. For users with motor limitations, having full voice control of their devices is truly transformative.

Voice Control offers an enhanced command and dictation experience. Users can traverse and control the entire screen with just their voices, giving them full access to every major function of the operating system. Additionally, users can gesture with their voices to click, swipe, and tap anywhere—so they can do everything someone could do with a mouse or with touch. Voice Control availability on macOS, iOS, and iPadOS ensures a consistent experience for users on all of their Apple devices.

Speech-to-text transcription

At the core of Voice Control is its ability to understand voices. By integrating the latest advances in machine learning for speech-to-text transcription, Voice Control is Apple's best built-in dictation technology yet. For users who can't type with their hands, accurate dictation is essential for fast and efficient communication. The speech recognition engine in Voice Control accurately understands natural speech so that users don't have to focus on saying a phrase perfectly.

By incorporating machine learning techniques focused on endpoint detection— or understanding when a user starts and finishes speaking—Voice Control differentiates between dictation and commands so that users can easily move between these two modes. For example, in Messages, if you say, "Happy birthday. Tap send.", only "Happy birthday" is sent, just as you intended. If you say, "Happy birthday. Delete that.", "Happy birthday" is transcribed and then deleted.

Voice Control settings include customization options in the Commands and Vocabulary tabs that make dictation even more powerful. Users can create custom words to communicate specialized terms for school or work. This is helpful when engaging in activities like writing a biology report, filling out a tax form, or explaining a technical concept. Users can also create custom commands to save time, such as "insert home address," to expedite the input of their addresses or "insert mobile" to add their phone numbers.

Text editing

Voice Control in U.S. English is available on iOS 13, iPadOS, and macOS Catalina and leverages the Siri speech recognition engine for accurate speech-to-text transcription. On macOS Catalina, Voice Control is also available in all 40 languages where Enhanced Dictation was previously available.

Voice Control builds on advanced dictation accuracy with a range of text editing commands that enable users to quickly make corrections and move on to expressing their next ideas. The main editing capabilities allow you to:

- Replace one phrase with another. For example, saying "Replace 'I'm almost there' with 'I just arrived'" will replace "I'm almost there" with "I just arrived."

- Position the cursor to make edits. For example, you can say, "Move up two lines. Move forward two words. Capitalize that." and Voice Control will capitalize the specific word you indicated in the paragraph. This eliminates the need to delete entire sentences and start again.

- Select text with precision. You can select the exact text you want, from single characters to an entire document. For instance, saying "Select previous word" will select the word right before the cursor, and "Extend selection backward by one sentence" will widen the selection to include the entire sentence.

- View word and emoji suggestions. For example, if you recently dictated the word "love" but meant to input a different word or even an emoji, you can say "Correct love," and a list of alternative words and emoji will appear. You can also insert emoji by name—for example, "Insert thumbs-up emoji" will insert 👍.
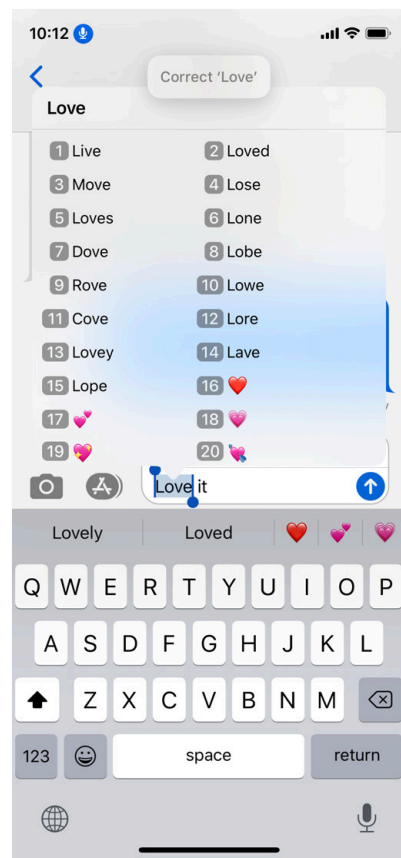
Voice command in Messages on iOS 13: "Correct love."

Comprehensive navigation

Voice Control gives users with motor limitations full and comprehensive access to the user interface (UI), so they can easily traverse the screen and accomplish complex actions with their voices, from dragging onscreen items to selecting unlabeled buttons. The tools that make every corner of the UI accessible include:

- **Navigation commands.** Users can quickly interact with the system and apps through common navigation commands using their voices. For example, users can say "Open Apple Pay," "Take screenshot," "Mute sound," "Save document," "Search for <item>" in Safari, or "Scroll up or down" in Apple News.

- **Item Numbers.** In situations where users don't have navigation commands, they can use a number overlay. Saying "Show numbers" assigns numbers to all clickable or tappable onscreen items, and users can then say a number to select the item they want. Item Numbers automatically appear in menus and are especially useful for selecting unlabeled buttons and disambiguating between a series of unnamed elements, such as photos.


Voice command in Photos on iOS: "Show numbers."

- **Item Names.** On iOS and iPadOS, Voice Control has the additional benefit of showing Item Names, which place a name next to each tappable item. Users can say "Show names" to view the accessibility labels for apps, files, buttons, and links, then say the name of the item they want to interact with. Developers can tag UI elements with Item Names and Item Numbers using the standard UIKit framework for views and buttons, which means users can have the same experience in both native and third-party apps.[1]
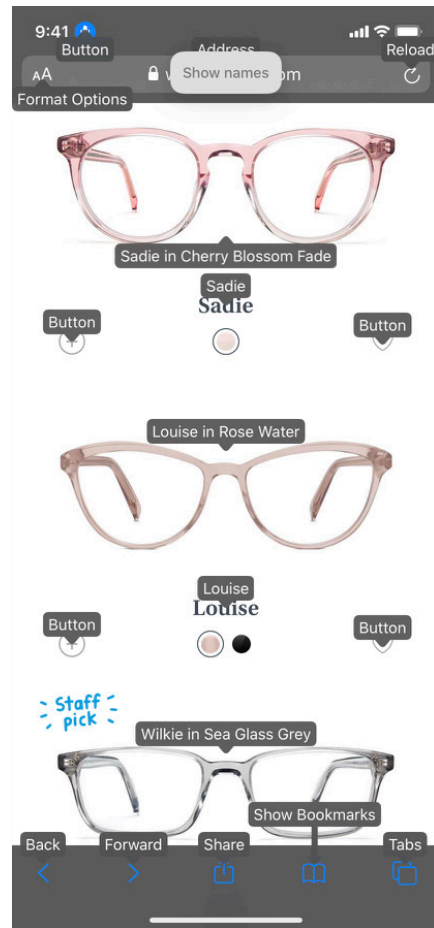
Voice command in Safari on iOS: "Show names."

- **Numbered Grid.** For unlabeled elements that are unreachable through Item Names and Item Numbers, users can use the grid overlay. Saying "Show grid" superimposes a grid with numbers on the screen, enabling users to iteratively drill into a box on the grid and interact with the item it contains. Numbered Grid provides fine-grain control to accomplish tasks like dragging an item to an unlabeled destination or dropping a pin in an undefined Maps location. With a grid overlay, users can also interact more deeply with apps that haven't fully incorporated accessibility labels.
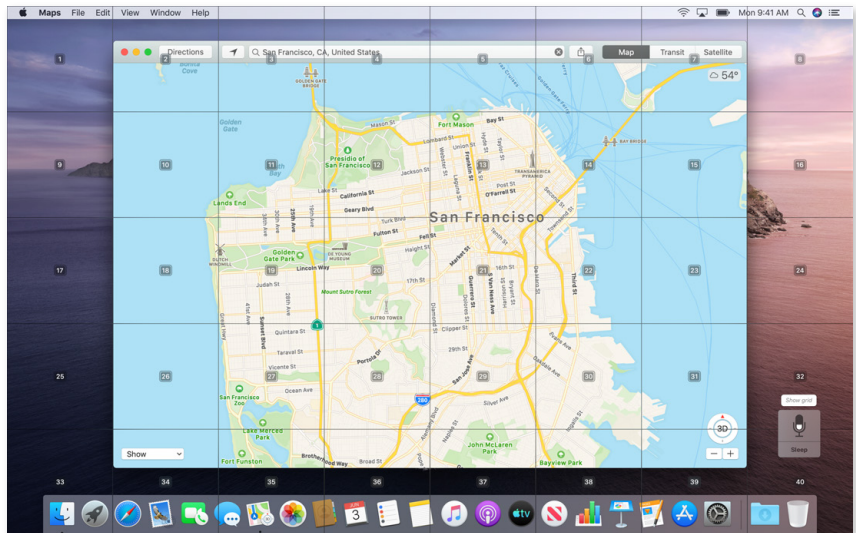
Voice command in Maps on macOS: "Show grid."

You can find a comprehensive list of Voice Control voice gestures for iOS 13 and iPadOS in Voice Control settings. Here are some examples:

• Swipe up

• Swipe to bottom

• Two finger swipe left at 5*

• Go home

• Double tap

• Tap and hold at 7*

• Two finger double tap at 7*

• Long press at 14*

• Scroll to bottom

• Scroll to left edge

• Pan left

• Two finger pan right

• Rotate clockwise

• Rotate to portrait

• Zoom in

• Decrease zoom

• Zoom right

• Start drag at 10*

• Drop at 20*

• Drag from 6 to 13*

• Cancel gesture

*The user has the Numbered Grid or Item Numbers turned on and is referring to specific numbers to move items across the screen.

• **Recorded commands.** On iOS and iPadOS, users can record a multistep process and give it a command name. For example, a user who frequently watches soccer on the Apple TV app could create a recorded command to quickly view what games are on. The user would begin by saying, "Start recording commands," then speak each step: "Open TV. Tap Sports. Scroll to bottom. Tap Soccer." Then the user would say, "Stop recording commands." A prompt would appear asking the user to name the new command—for example, "Browse soccer games." From then on, if the user says, "Browse soccer games," the Apple TV app will automatically launch a view that shows live and upcoming soccer games.

Voice gestures

On iPhone and iPad, Voice Control enables users to perform Multi-Touch gestures like tap, double tap, and scroll up or down with their voices to fully navigate the operating system without touching their devices. Users can also record Custom Gestures. For example, an avid gamer could create commands to jump, swipe, or tap specific areas onscreen. After saying, "Create new command" in Voice Control settings, the user would say "Action," then "Run Custom Gesture" to open a recording screen. The user could keep the Numbered Grid on while saying "Drag <number> to <number>" to create a jumping gesture, then say "Tap stop." After naming the gesture—for example, "Jump up"—the user can say this name to enact the gesture while playing the game.

Attention awareness

On iPhone and iPad models with the TrueDepth camera, Voice Control intelligently activates and deactivates depending on where the user is looking. The TrueDepth camera projects and analyzes over 30,000 invisible dots to create a depth map of the user's face and also captures an infrared image of the face. In addition to being used for Face ID, this data is also used to create the Attention Aware function, which recognizes when users' eyes are open and their attention is directed toward the device. With Attention Aware turned on in Voice Control settings, Voice Control goes to sleep when users look away from the camera and wakes up when users look toward the camera, enabling them to easily move between interacting with their devices and with people around them.[2]

On-device processing for privacy

Apple believes that privacy should be equally accessible to all users, however they interact with their devices. By leveraging the processing power of Apple's A-series chips on iPhone and iPad and the unique silicon architecture of Mac, Voice Control audio processing happens on-device while maintaining fast performance. This keeps the words you use to control your devices private, from the messages you dictate and the news stories you tap to the websites you scroll through.

An additional benefit of on-device processing is that Voice Control will always function. Even if you're out of cellular range or Wi-Fi is down, you have complete control of your device and can continue engaging in locally based activities like writing, coding, editing images, and listening to downloaded content.

Voice Control on macOS, iOS, and iPadOS

The new Voice Control on macOS Catalina, iOS 13, and iPadOS vastly expands what users can achieve with their voices, from reaching the furthest corners of the UI to engaging in apps more deeply than ever before. The cross-platform availability of Voice Control provides a consistent experience across Mac, iPhone, and iPad, and on-device processing enables users to always have access to their devices. For users with motor limitations, this powerful built-in tool transforms how they work, play, create, and connect on Apple devices.

What's the difference between Voice Control and Siri?
Voice Control lets users control the entire device with spoken commands and specialized tools, while Siri is an intelligent assistant that lets users ask for information and complete everyday tasks using natural language. Voice Control offers comprehensive capabilities such as voice gestures, name and number labels, grid overlays, text editing commands, and deep customization, while Siri assists with setting reminders, making appointments, looking up directions, and learning game scores.

Can you use Voice Control and Siri at the same time?

Absolutely. For example, after setting up "Hey Siri" on iOS, a user can say, "Hey Siri, navigate me home," and Siri will launch directions in Maps. Then the user can use Voice Control commands like "zoom in" to interact with the map.

Can anyone use Voice Control?

Anyone can learn to use Voice Control. Some users might want to use just the dictation and editing elements of Voice Control, formerly known as Enhanced Dictation on macOS, while others will want to use all Voice Control features.

What if I just want to control my device and not use dictation?

Users can say "Command Mode" to instruct Voice Control to ignore dictation and respond only to commands, and they can say "Dictation Mode" to instruct Voice Control to listen for both dictation and commands.